Experiences of open science & software

Peter Bankhead

Senior Lecturer in Digital Pathology

University of Edinburgh

11/2018

What is Bioimage analysis?

Biology

Not what I do

Image analysis

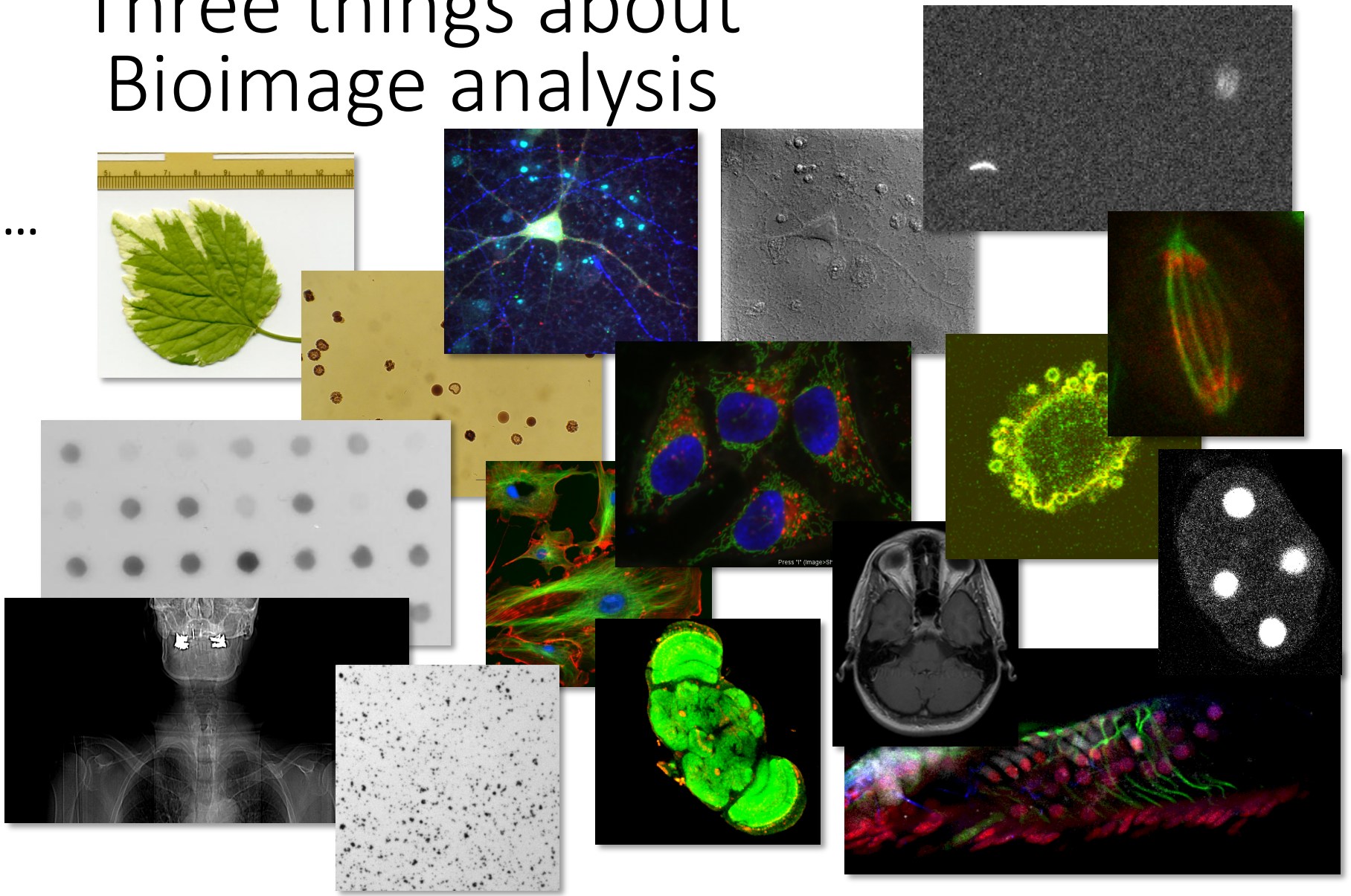
What I do

Three things about Bioimage analysis

Three things about Bioimage analysis

Bioimage analysis…

- …is diverse

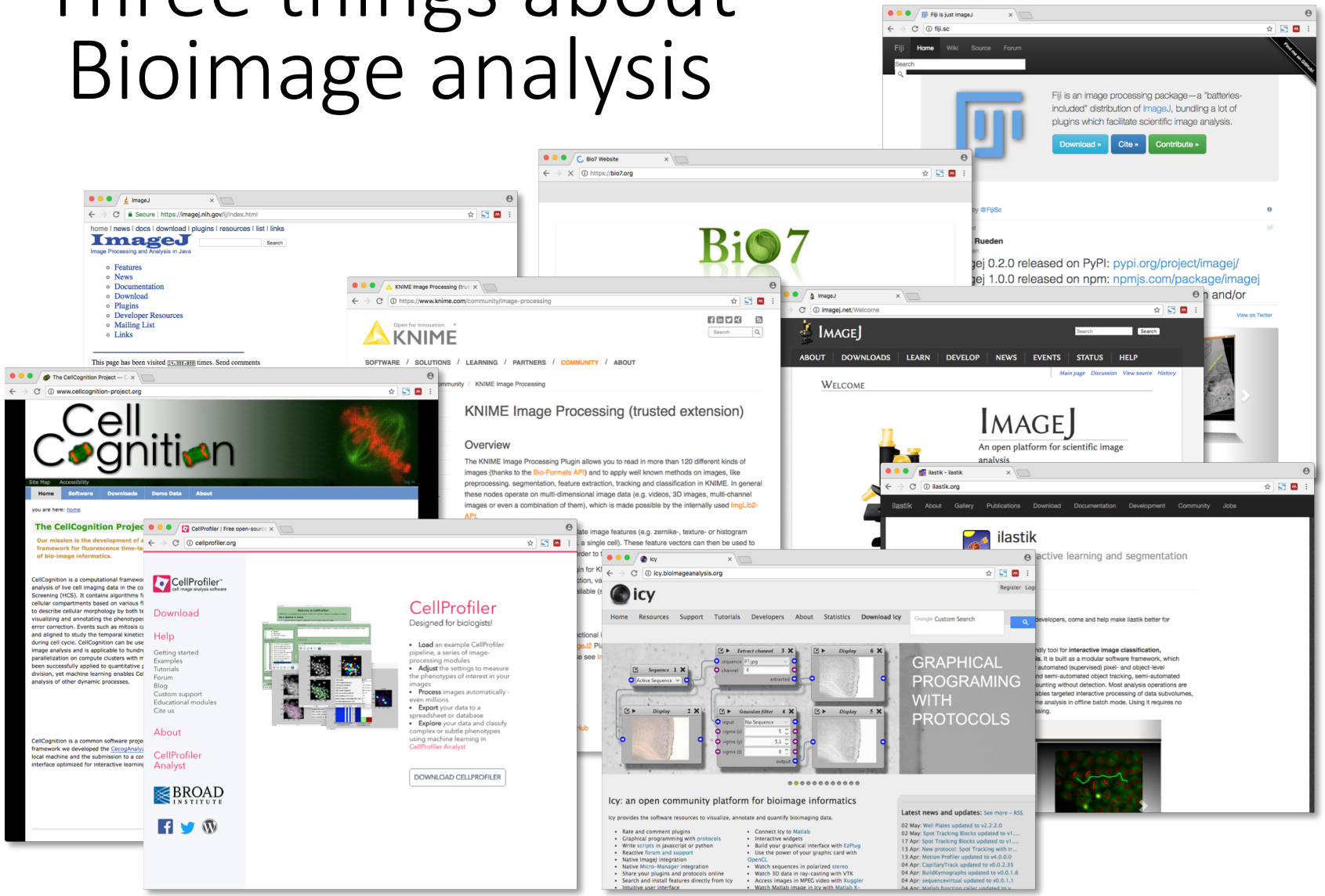


Sample images from *ImageJ* & *Fiji*

Three things about Bioimage analysis

Bioimage analysis…

- …is diverse



Popular open source bioimage analysis software

Three things about Bioimage analysis

Bioimage analysis…

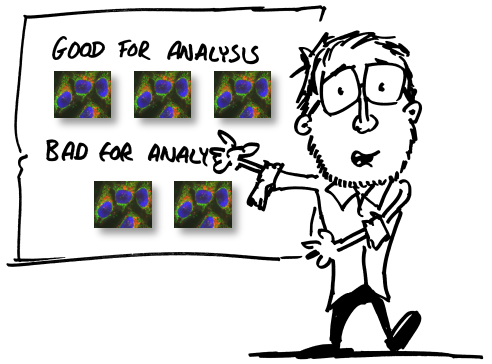
- …is diverse
- …is cross-disciplinary

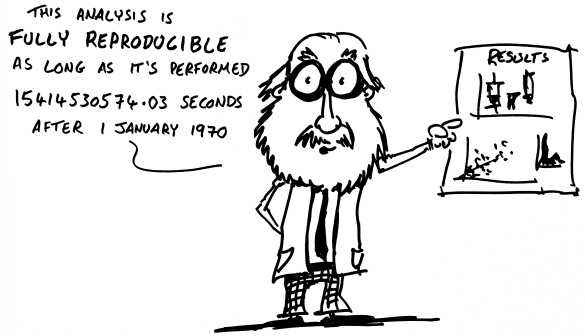
Sometimes the same person masters both skillsets – but more often we need to work together

Three things about Bioimage analysis

Bioimage analysis…

- …is diverse
- …is cross-disciplinary
- …is often misunderstood

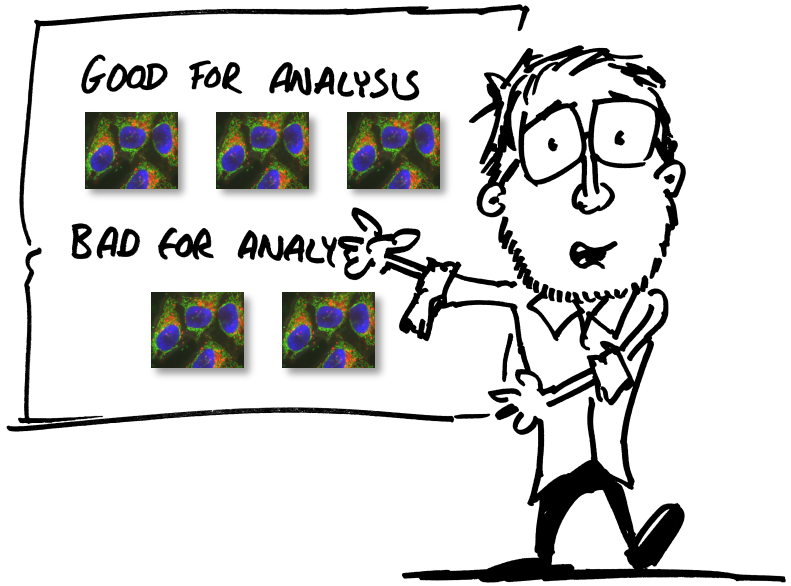




Three things about Bioimage analysis

Bioimage analysis…

- …is diverse
- …is cross-disciplinary
- …is often misunderstood

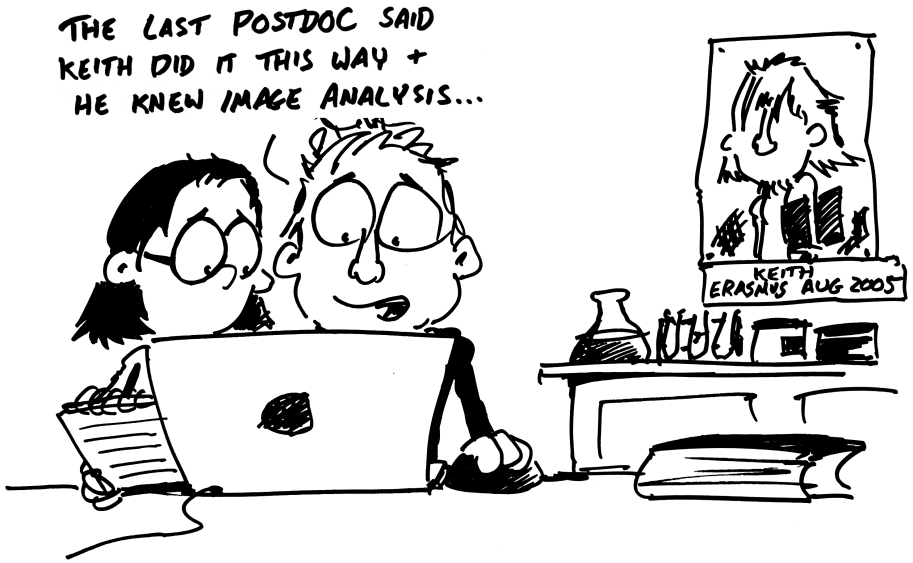


Not all images are suitable for quantitative analysis

Three things about Bioimage analysis

Bioimage analysis…

- …is diverse
- …is cross-disciplinary
- …is often misunderstood



Just because someone used a method before doesn’t mean it’s right

Three things about Bioimage analysis

Bioimage analysis…

- …is diverse
- …is cross-disciplinary
- …is often misunderstood

Image analysis is *not* unbiased
(even if a computer does it)

Three things about Bioimage analysis

Bioimage analysis…

- …is diverse
- …is cross-disciplinary
- …is often misunderstood

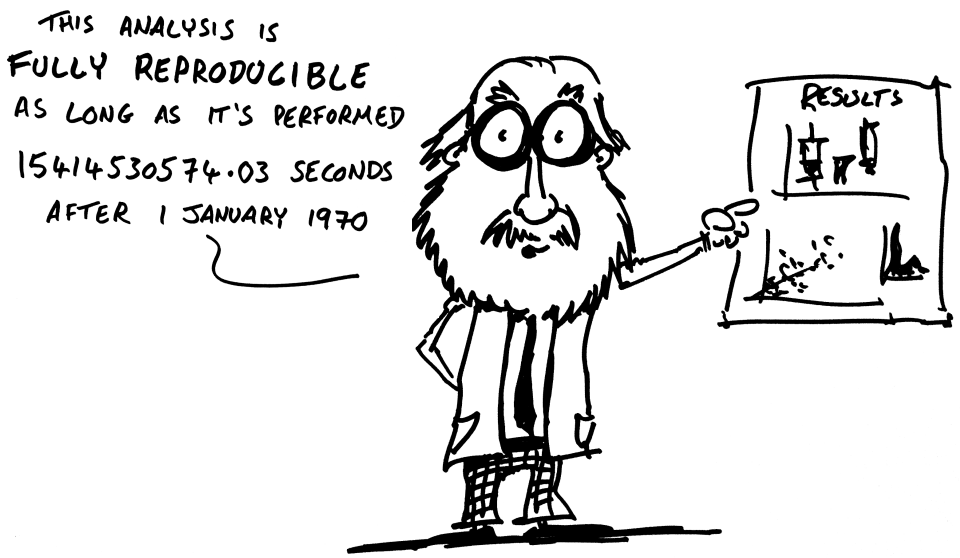
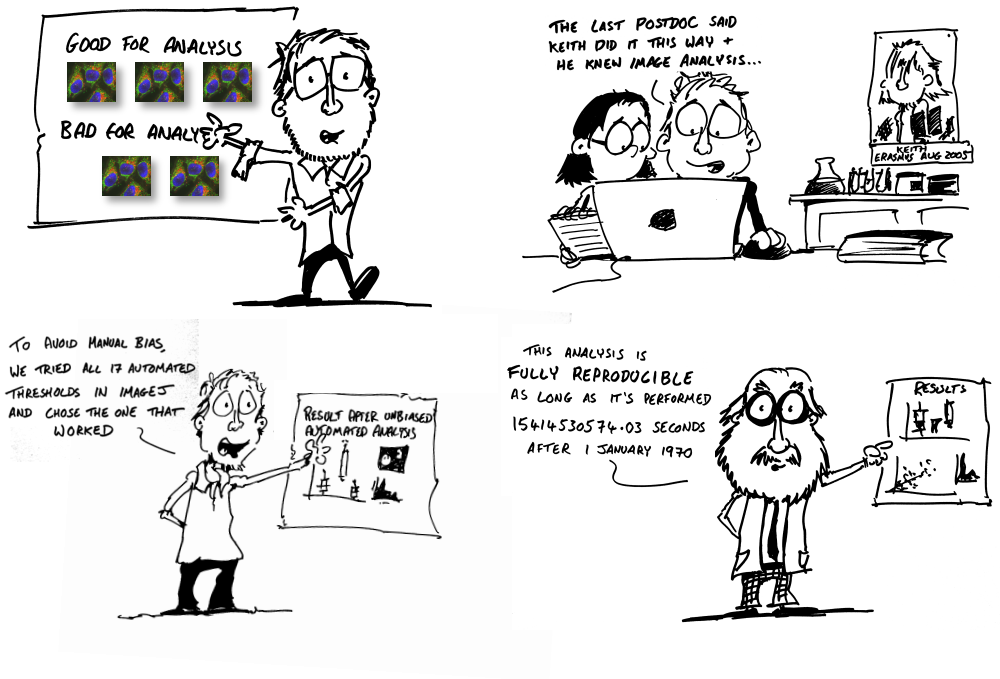


Image analysis *should* be reproducible –
but *it might not be!*

Open data & methods are essential!

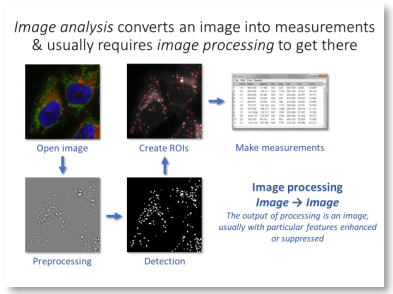
We need to know *exactly*:

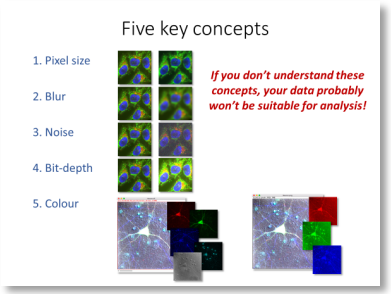
- What was analysed — *Image repositories*
- How it was analysed — *Open software*
- What were the results — *Numbers & visualizations*

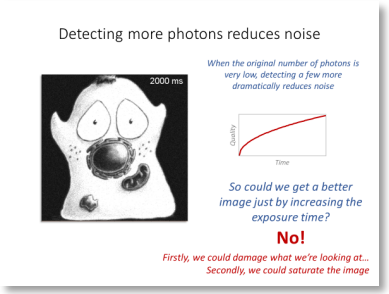


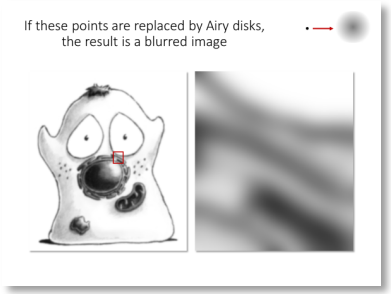
Not every assumption will be recognized or reported –
we must be able to go to the data & analyse in a different way

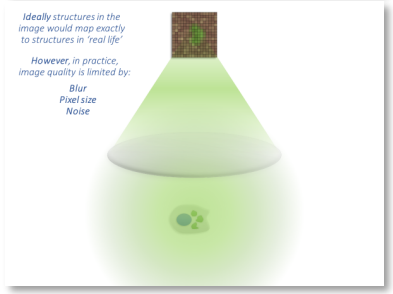
In Heidelberg, I chose to focus on teaching

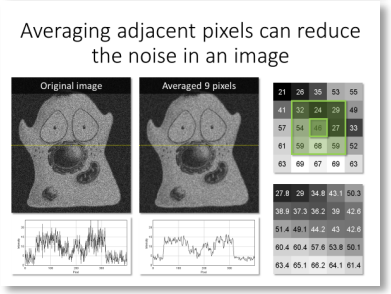


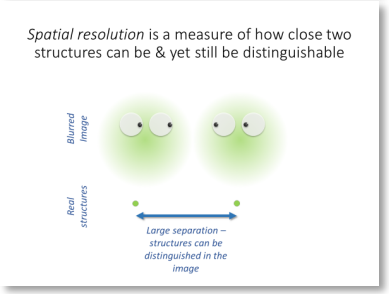


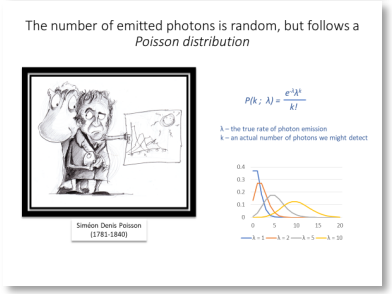


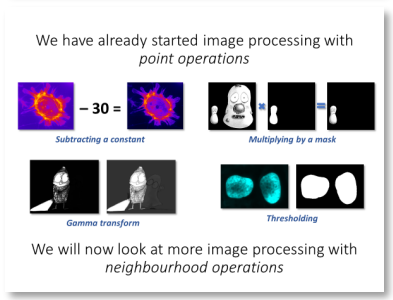


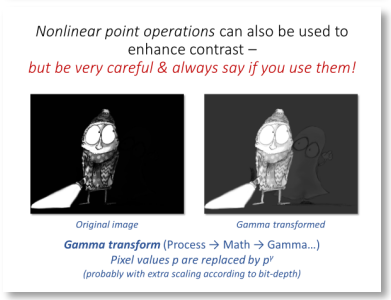


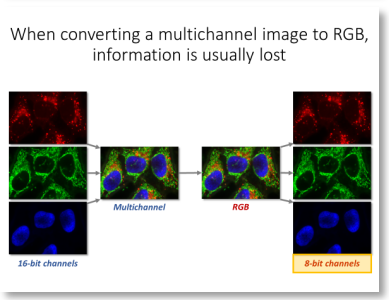


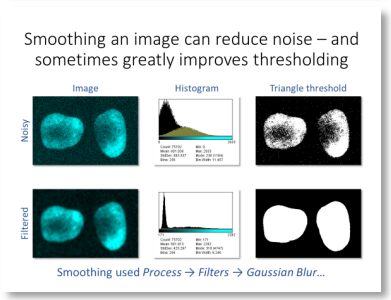


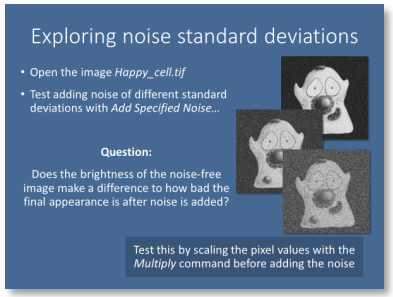


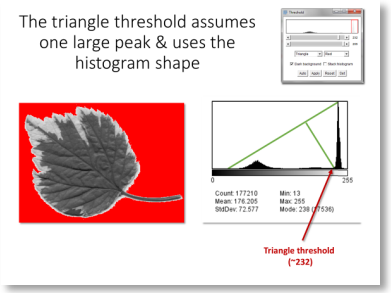


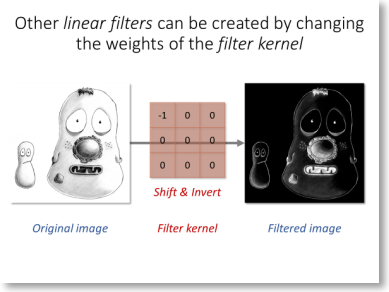


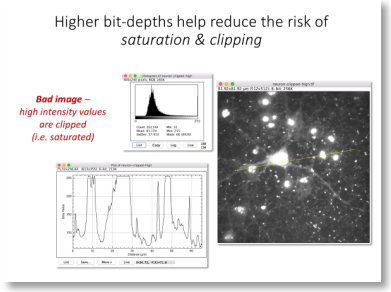


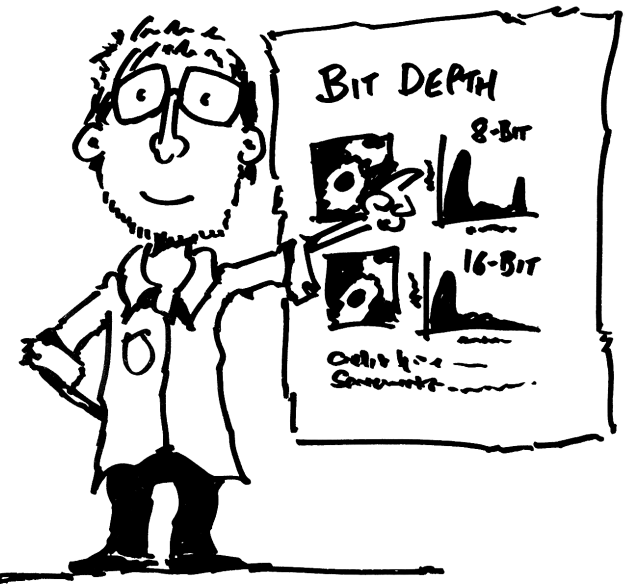






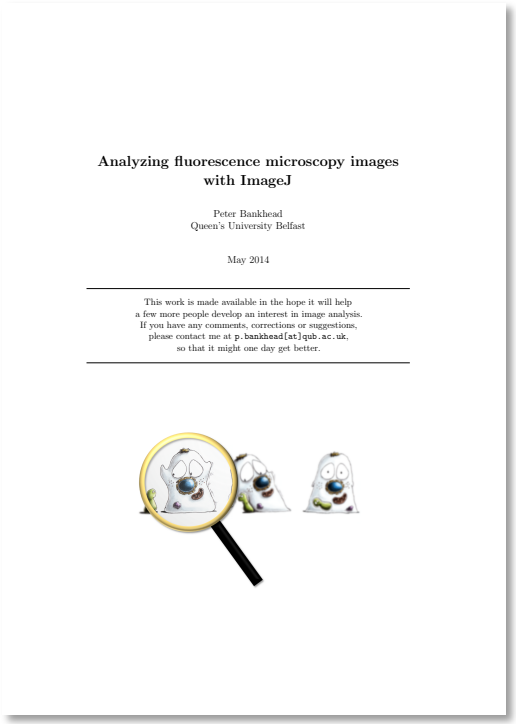




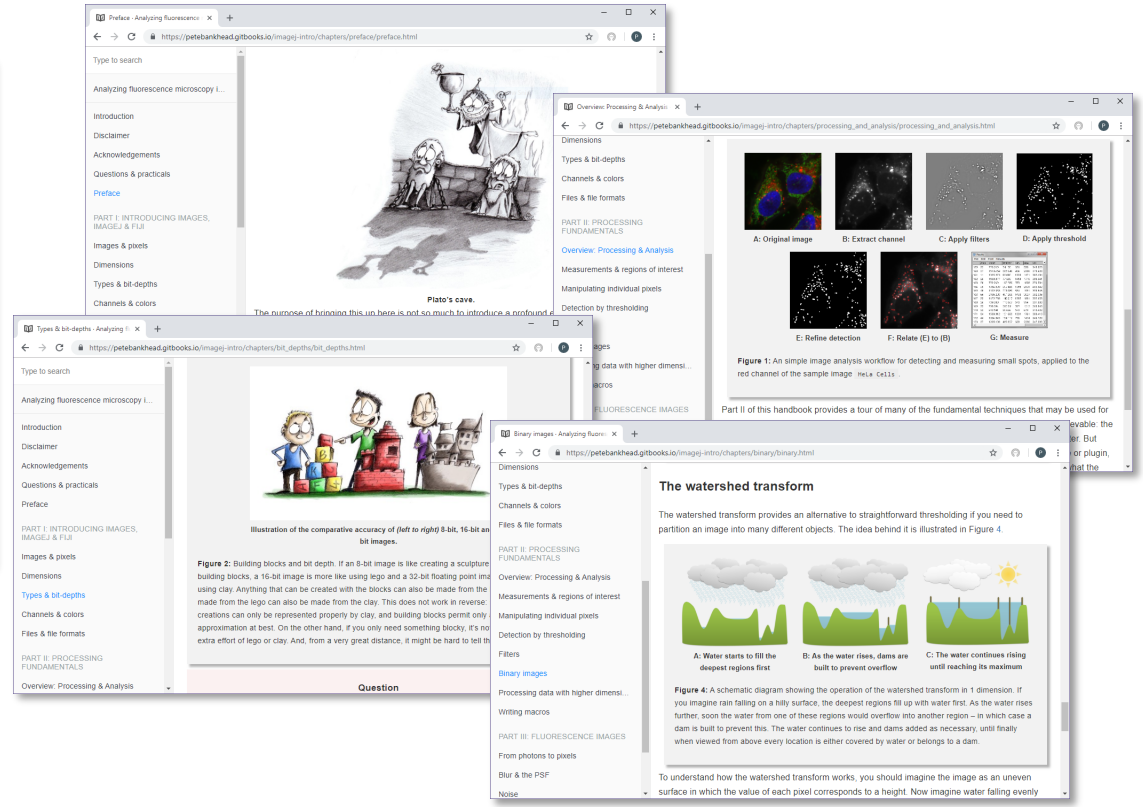


This was useful,
but its impact was limited
by time & proximity…

Analyzing fluorescence microscopy images with ImageJ



PDF version



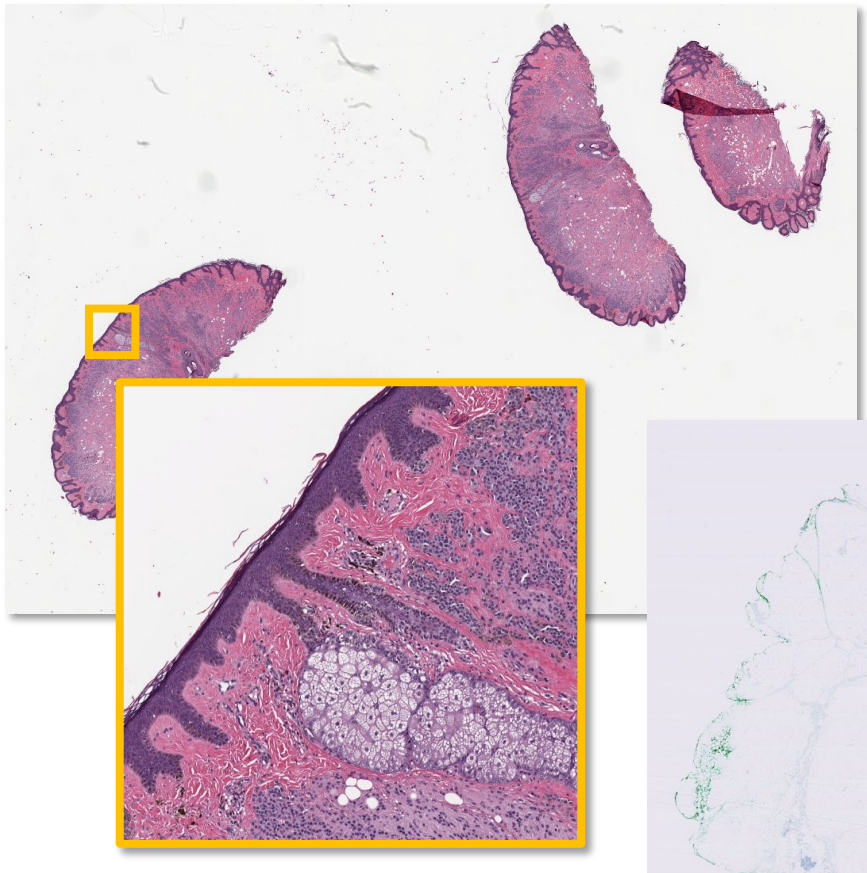
GitBook version

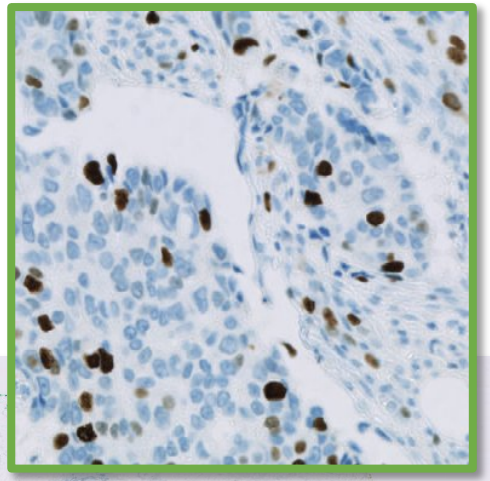
https://petebankhead.gitbooks.io/imagej-intro

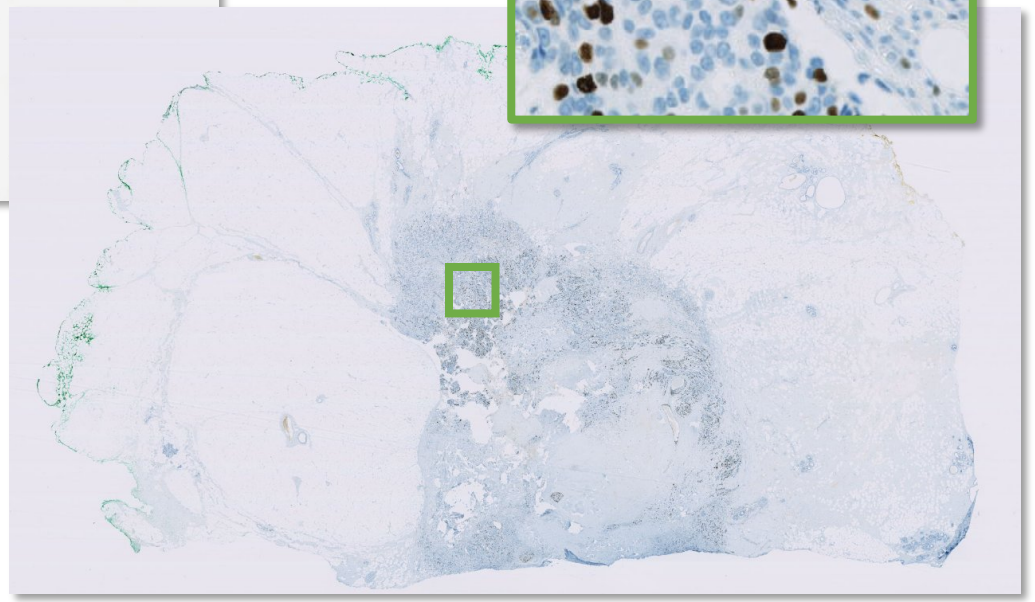
New kinds of images require new kinds of software

Sometimes the necessary tools just don’t exist

Pathology bridges science & medicine and plays a crucial role in patient care

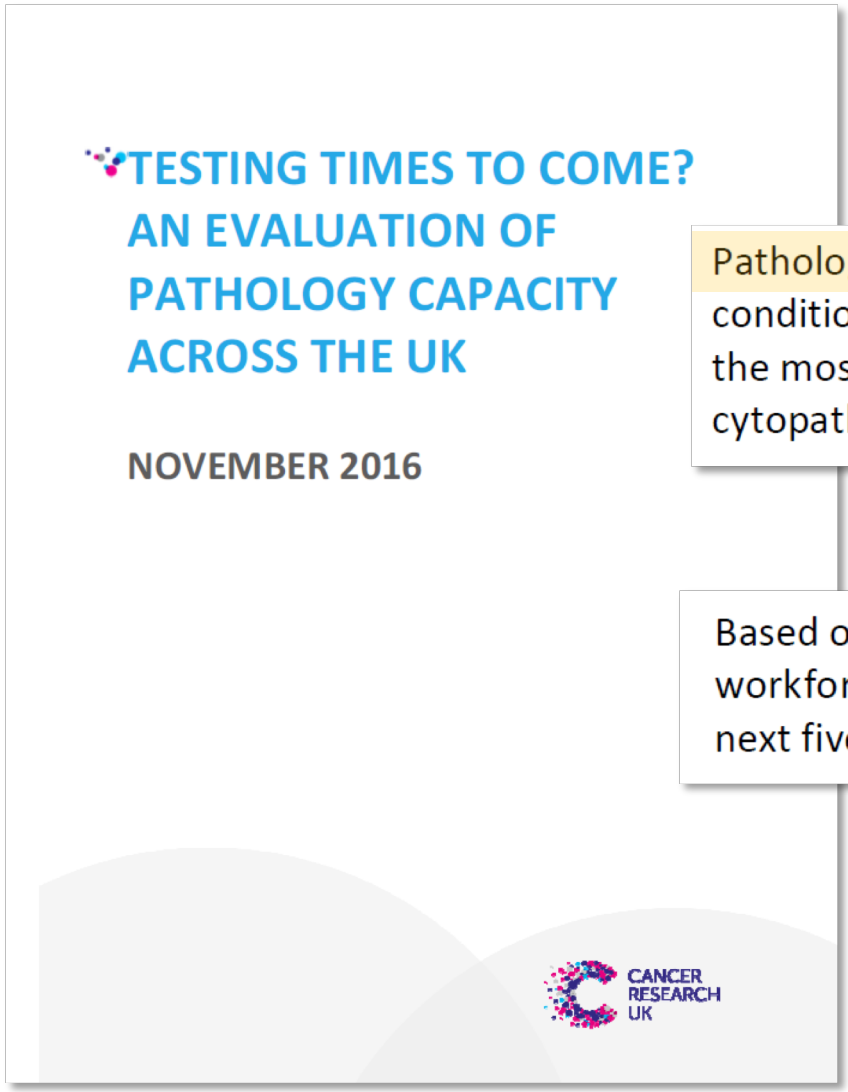






Sample images from *https://openslide.org*

There is a critical shortage of pathologists

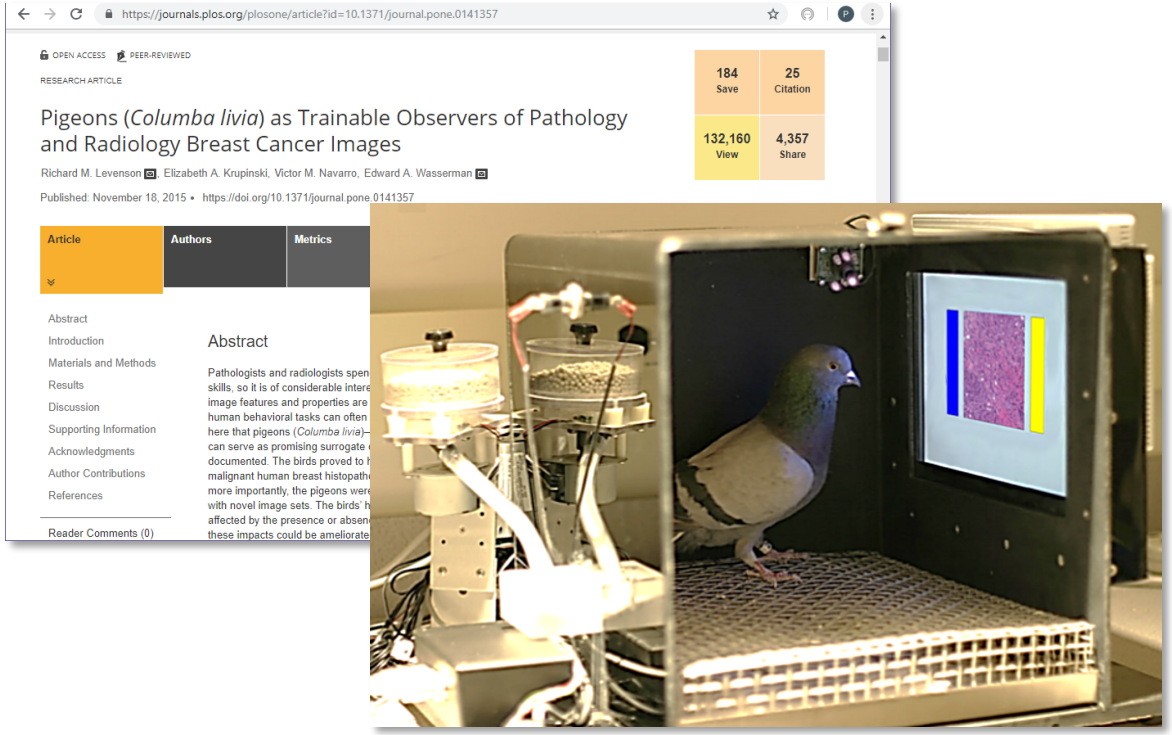


Pathology plays a major role in the diagnosis and treatment of cancer, as well as many other conditions. Pathology is comprised of 19 different disciplines and our research focussed on the most relevant to cancer: cellular pathology (which encompasses both histopathology and cytopathology); blood sciences; and molecular pathology.

Based on the number of pathologists currently in training and the age profile of the current workforce, our study found there is likely to be a severe crisis in pathology capacity within the next five to ten years.

There are two main approaches to address the needs in pathology

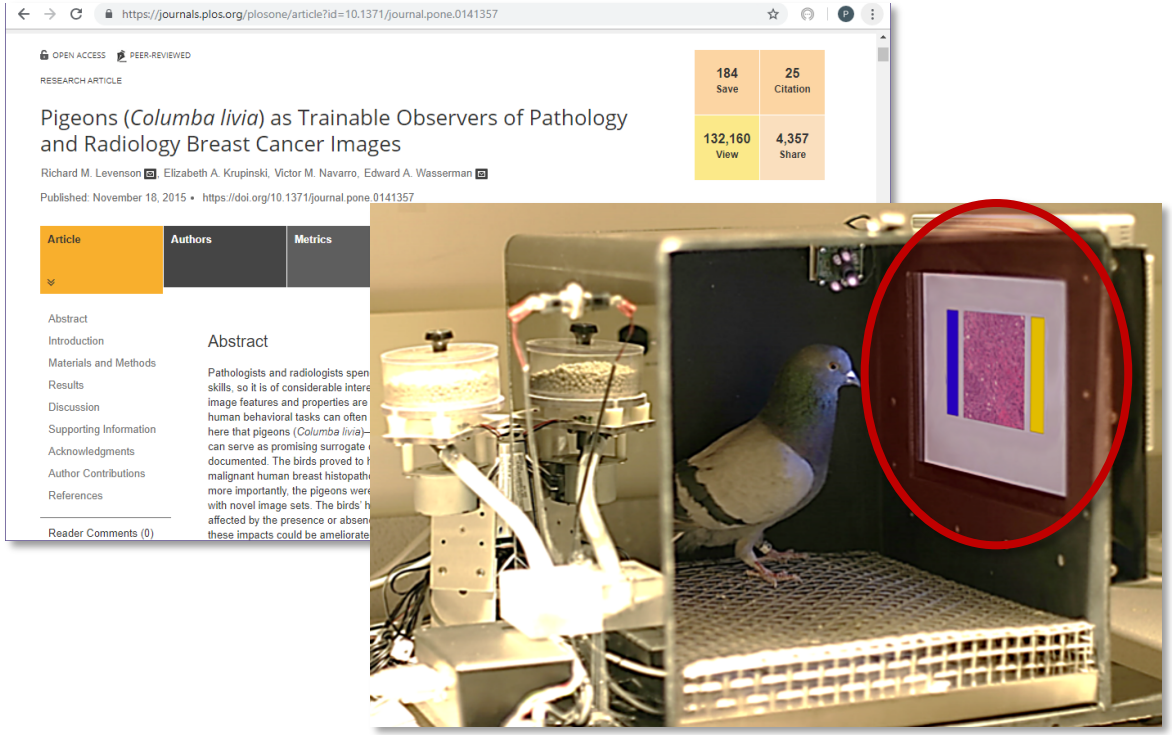
Digital pathology



Levenson RM, *et al*. PLoS ONE (2015)

Pigeons

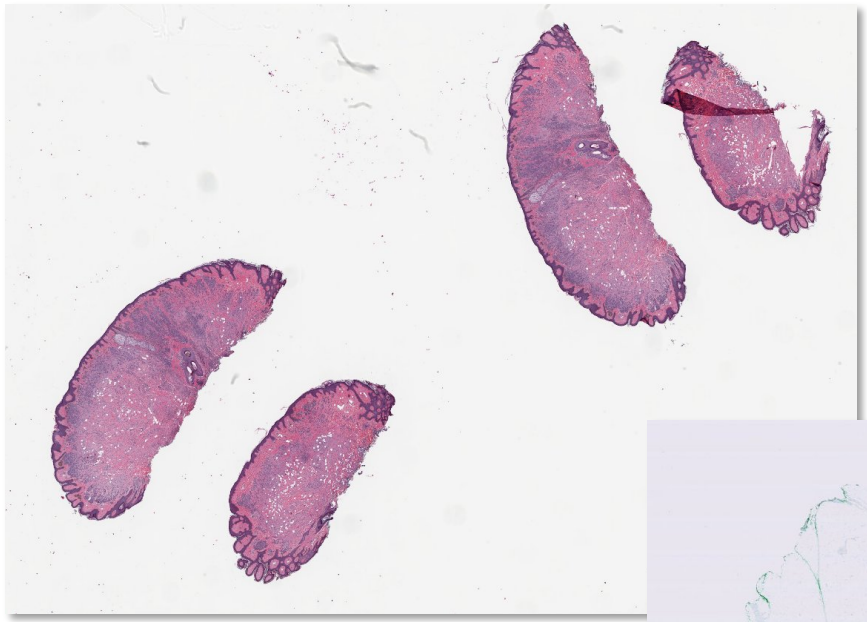
There is *really only one* main approach to address the needs in pathology



Levenson RM, *et al*. PLoS ONE (2015)

Digital pathology

Digital pathology is based on *whole slide images*



Whole slide images are very big & very complex

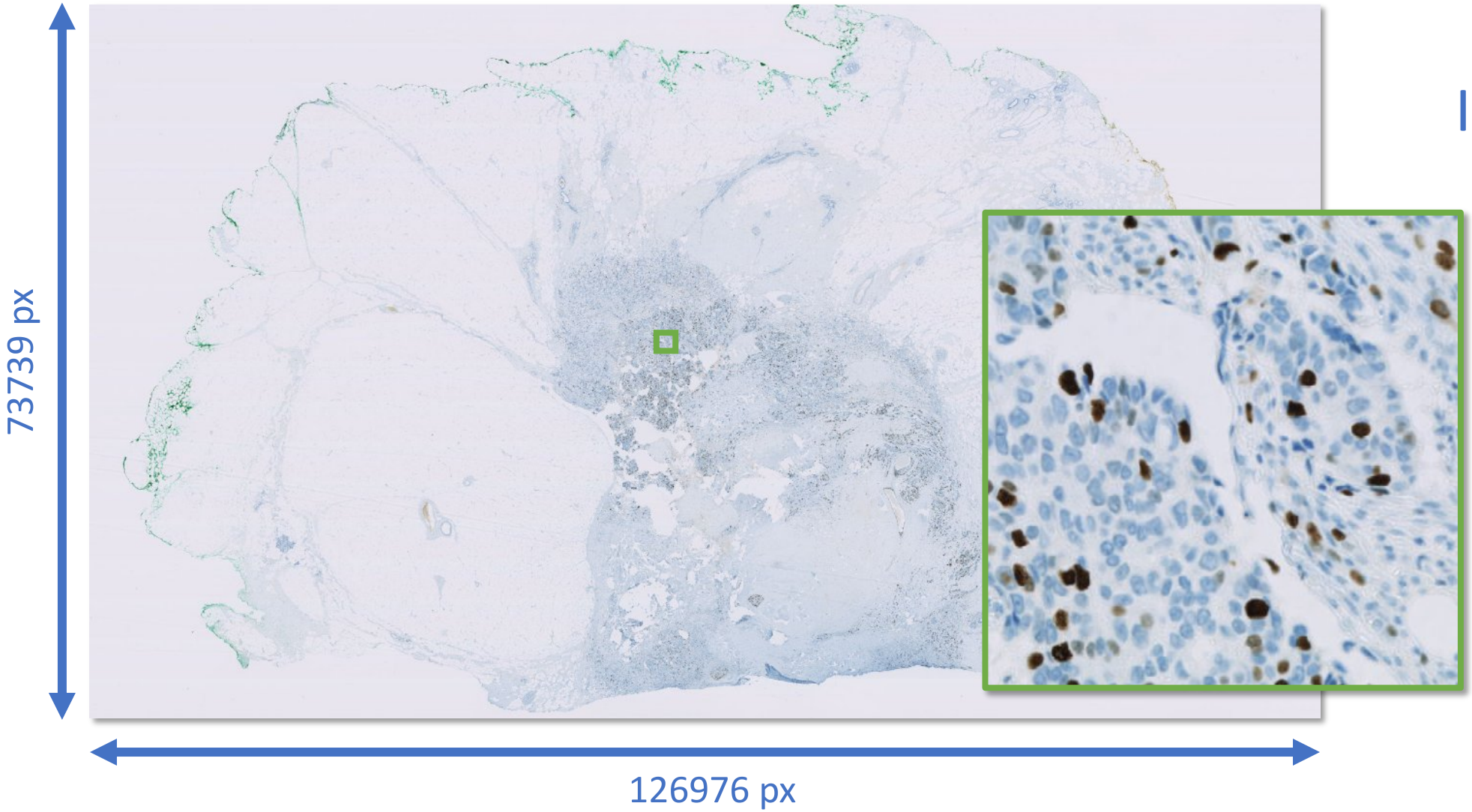
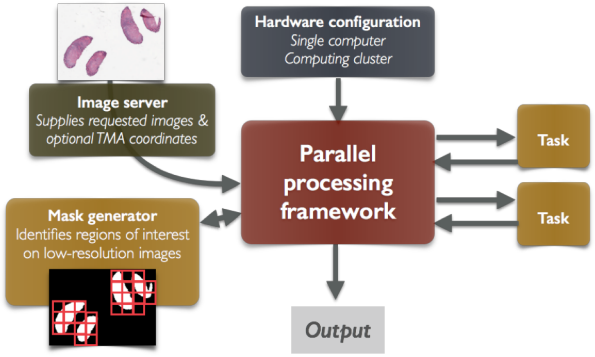
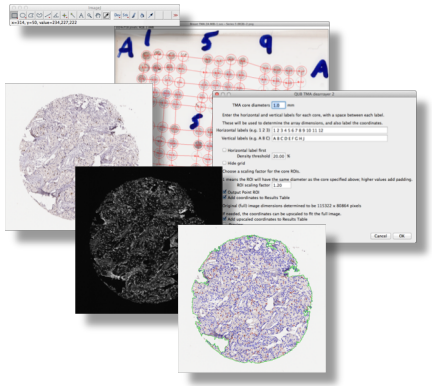


Image size:
35 GB

Most bioimage analysis tools don’t support whole slide images



Attempt #1
Python scripts + MPI

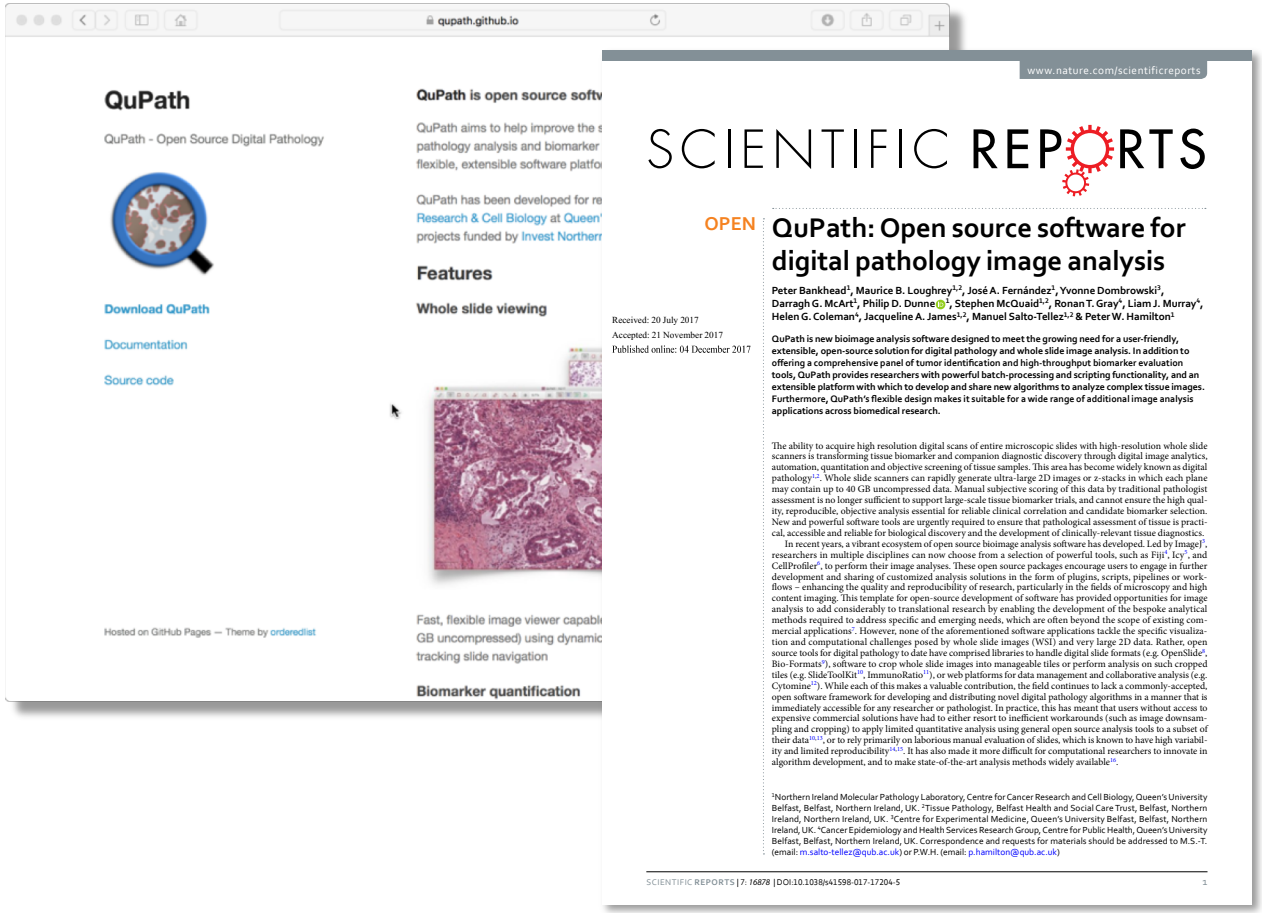


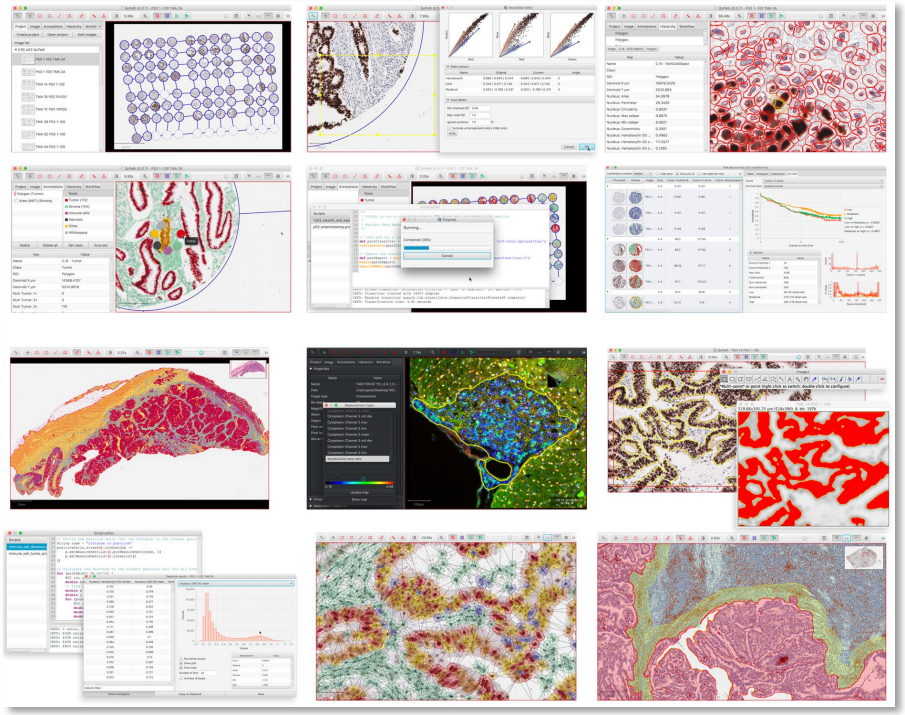
Attempt #2
ImageJ plugins

Without the right tools, analysis methods are slow to create, hard to apply, not transferable… and not very useful

QuPath:

Open source, whole slide image analysis software





Some lessons learned…

There's quite a lot involved in creating a new software platform…

Interactive whole slide image viewer

Interactive drawing tools

Scanner image format support

Region of interest representation

Image caching

Dearraying Tissue Microarrays

Object representation

Cell segmentation

Feature computation

Quantification of nuclear, cytoplasmic & membranous staining

Tumour identification

Machine learning

Visualization of results

Membrane detection

Texture computation

Data management

Stain estimation

Digital stain separation

User interface design

Efficient data structures

Command logging

Help documentation

Image tiling

Automation

Parallel processing

Scripting support

Statistical analysis

Import / Export

Multidimensional data support

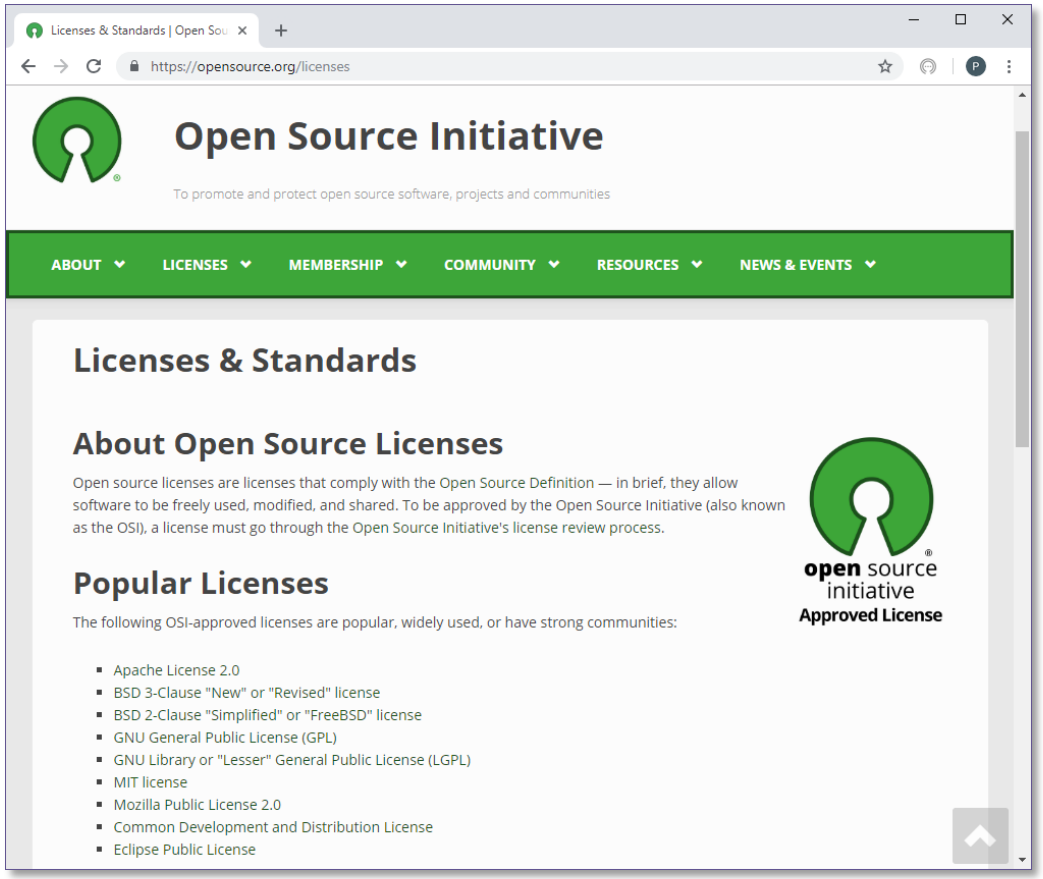
Integration of survival data

A research paper could focus on one or two of these

Open source can be scary

Open source can be complicated

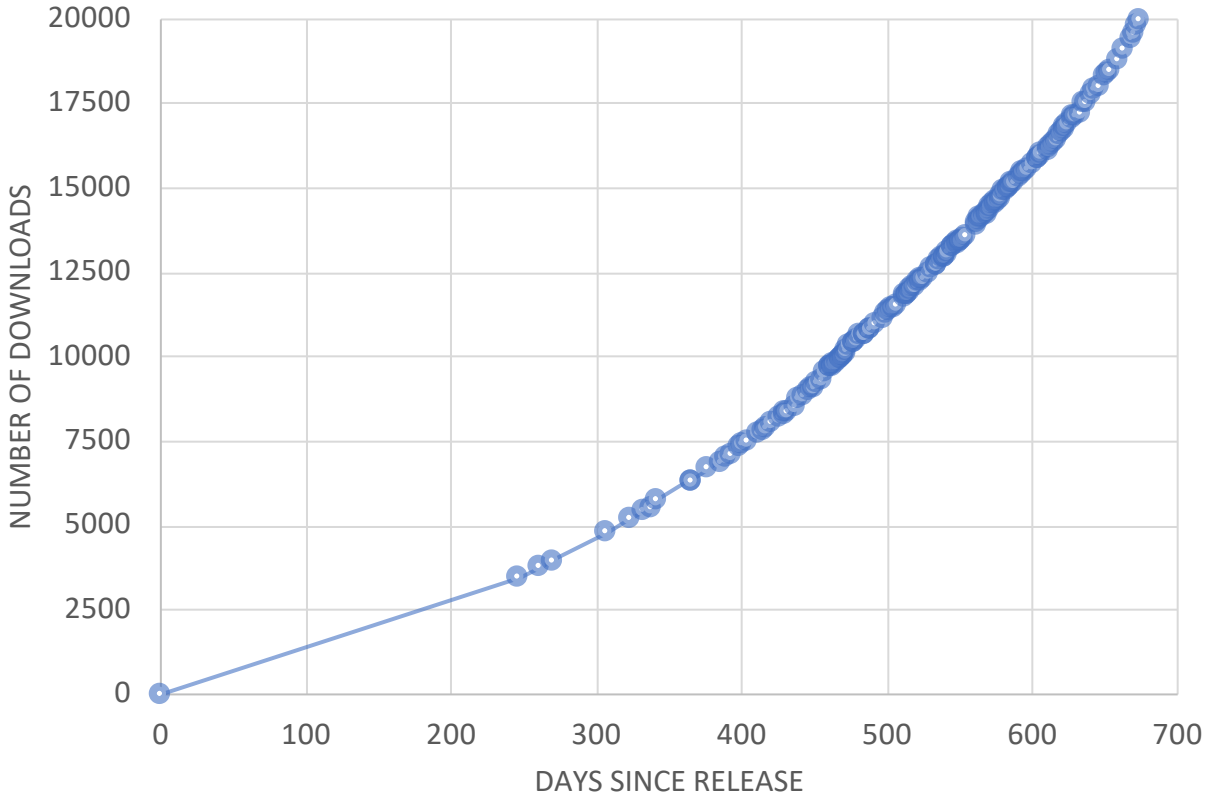
The license choice matters

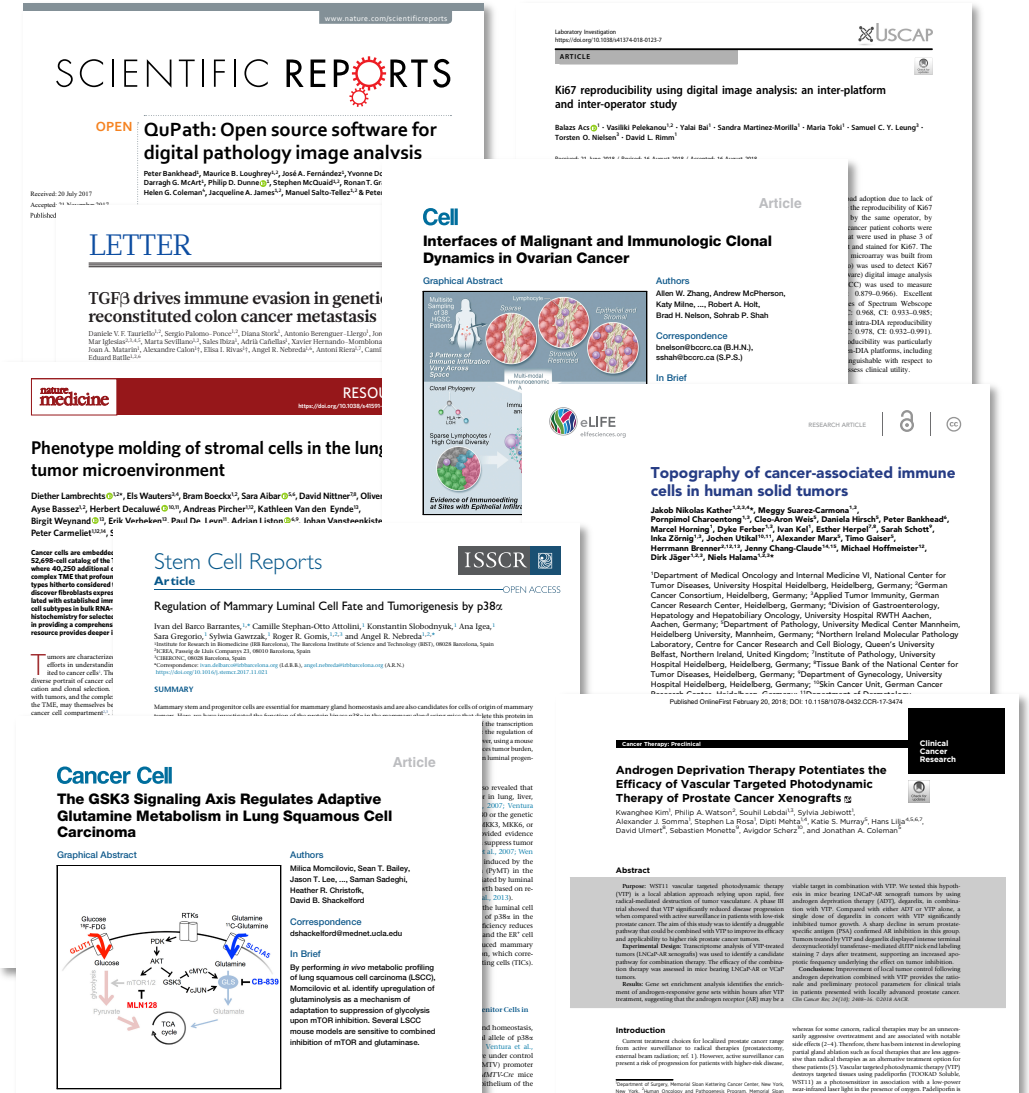


https://opensource.org/licenses

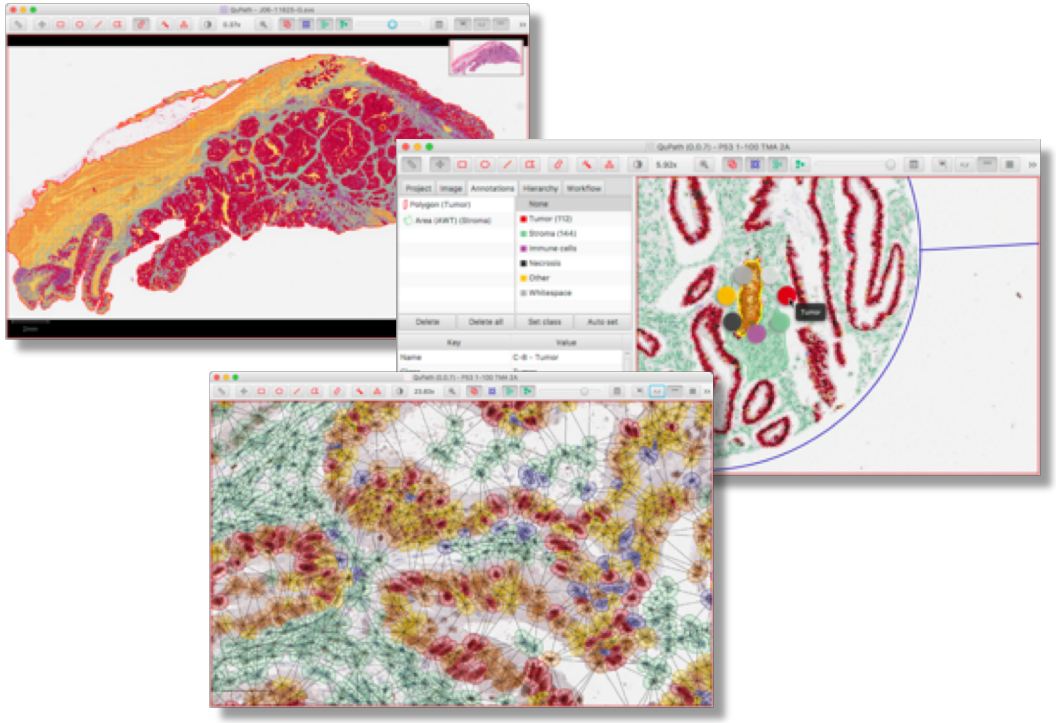
Open source can be costly

Open source can be transformative

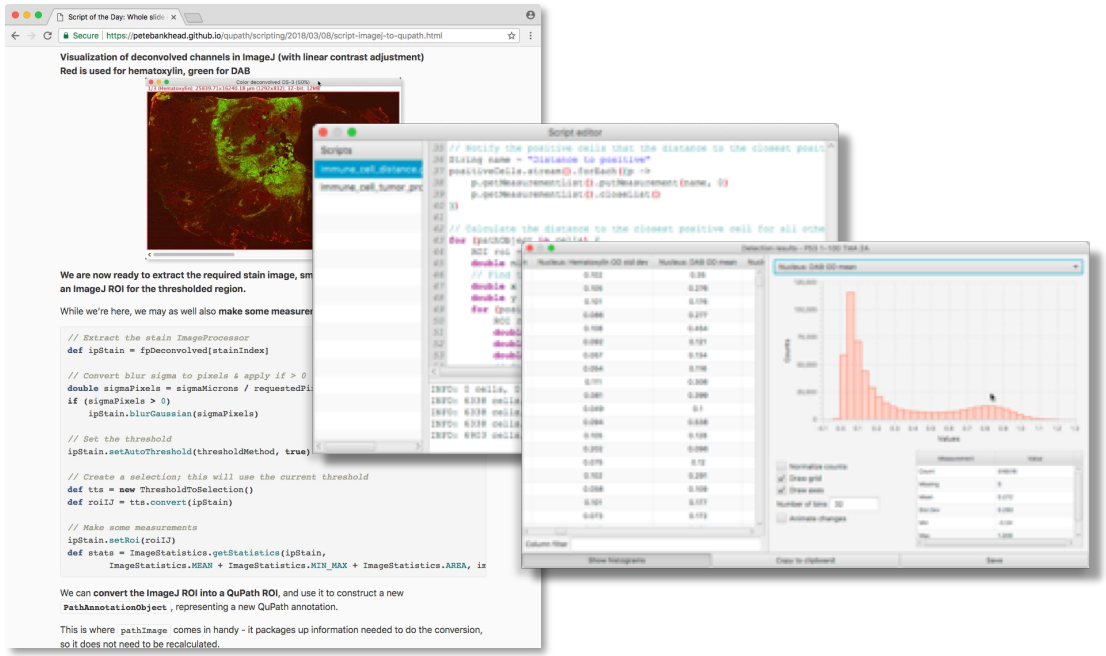




User-friendliness & developer-friendliness are *both* important



Visualizations are key for cross-disciplinary collaboration

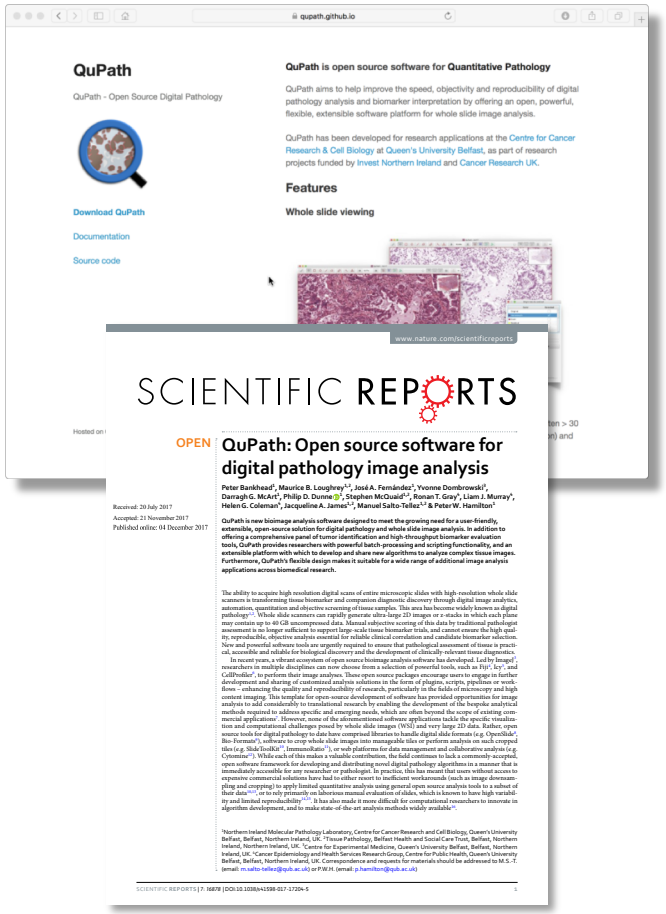


Extensibility & scripting support make far more things possible

Reflections & Opinions

- Developing & maintaining new software can take a long time
  - But the impact it makes can *vastly* outweigh the time & cost of development
- Creating software in academia is harder than it should be
  - 'Is it really *research*? Shouldn't companies be doing it? Should we start a company…?'
  - If you use research software, *please at least cite the papers!*
- Open science is difficult & scary in practice… but absolutely necessary
  - *If* bioimaging data & analyses were fully available & transparent, I suspect we'd find
    - *Lots of poor-quality data ends up in high-quality journals*
    - *Analysis is often wrong / answers a subtly different question from what the authors thought*
    - *There are many new insights to be had using existing data*

Thanks for listening!



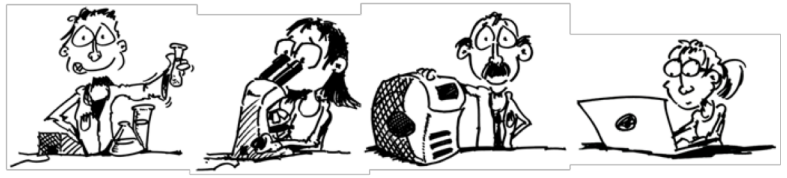
QuPath is software for **whole slide image analysis**...

...that is **freely available & pathologist-friendly** ...

...and **isn't locked down** with fixed algorithms...

...but is **open source, extensible & scriptable** for developers...

...and has a **wide user base** sharing ideas & insights...



https://qupath.github.io/